Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

STANDARDIZED TESTS AND THE IMPROVEMENT OF TEACHING IN FIRST-YEAR ALGEBRA[1]

H. O. RUGG
School of Education, University of Chicago
AND
J. R. CLARK
Parker High School, Chicago

A. NECESSARY CRITERIA FOR A PROGRAM FOR THE IMPROVEMENT OF INSTRUCTION

It is possible to attack the problem of the improvement of instruction in a high-school subject in a scientific experimental way. To do so in the most complete fashion will necessitate the setting up of a thoroughgoing program embracing the following steps:

1. Stating definitely the aims and outcomes of instruction in the particular course of study.
2. Classifying clearly the subject-matter of the course on a basis of the principal modes of learning involved in its mastery; a statement of the content of the course, from a standpoint both of the amount of material included and of the classification, arrangement, and order of presentation, supplemented by a detailed analysis of the mental processes called into play by different types of subject-matter.
3. *Designing and giving tests which will adequately measure ability in each of the fundamental phases of the subject-matter agreed upon.*
4. Critically evaluating the testing so as to give a complete and differentiated statement of fundamental weaknesses in learning (e.g., as revealed by the typical errors made by pupils).
5. Setting up experimental attempts to eliminate these fundamental weaknesses; from the standpoint of economy of time this

---

[1] A final report on the "Experimental Determination of Standards in First Year Algebra," made to the Mathematics Section of the University of Illinois High-School Conference, November 24, 1916.

means *the design of practice exercises, the determination of specific "best ways" of presenting material, of order of presentation, and of optimum length of drills.*

In brief, the writers have carried through to completion the first four steps of this program and have begun work on the fifth step. We believe that sufficiently adequate standards have been set up to enable teachers to attack the problems of teaching first-year algebra in a thoroughly experimental way. We suggest careful consideration of the results of the work to date, and the thoroughgoing co-operation of teachers of algebra generally, in improving instruction in the subject.

I. THE AIMS AND OUTCOMES OF INSTRUCTION

The first step in the program of improving teaching in first-year algebra involves the definite statement of the aims and outcomes of instruction. In the preliminary report of this investigation (see *School Review*, January, 1916) we made a complete statement of these aims and outcomes as follows:

a) THE AIM OF FIRST-YEAR ALGEBRA INSTRUCTION

The conduct of this investigation is based upon the following statement of the ultimate aim of first-year algebra: the successful use of algebraic symbolism in meeting new problem-situations, either (1) purely algebraic situations; (2) general mathematical situations (such as are found in mathematics courses other than algebra); or (3) non-symbolic situations, including various types of "practical" or "applied" problems not expressed in mathematical language.

b) THE OUTCOMES OF FIRST-YEAR ALGEBRA

Believing that some statement of the assumed outcomes of first-year algebra should be made the basis for further analysis, the following tentative statement of these outcomes is made:

A. *Immediate, specific, and preparatory outcomes.*—These include the comprehension, interpretation, and manipulation of the specific "automatic" operations involved in algebraic solutions; e.g., the four fundamentals as used with various type problems, factoring, removal of parentheses, etc. Certain of these are to be used in many specific algebra problems and in the solution of other mathematical problems; they are to be mastered as tools, preparatory to the taking of other mathematical courses, as well as for use as automatic tools for the solution of all types of "applied" problems. These involve primarily the learning of rules and the formation of specific habits of manipulation.

B. *Immediate generalized outcomes.*—These involve recourse to selective, analytic, and conceptualizing abilities; ability to apply principles, in addition to ability to remember rules and to make certain fundamental habitual adjustments in the solution of general practical and applied problems and in solution of problems belonging primarily to other mathematical fields.

C. *Remote and less tangible outcomes.*—The development of the ability to deal with general number concepts and "think quantitatively"; the development of attitudes of (1) orientation in algebraic or general mathematical fields containing problem-situations; (2) confidence in one's ability to use successfully algebraic symbols in meeting new situations; (3) a broadened intellectual background or perspective for the general cultural comprehension and interpretation of the scientific methods by which technical problems may be solved, i.e., the development of a "scientific attitude." In the present attack on the general problem of standards in this field it is not proposed to consider this third group of outcomes. Standards in first-year algebra, we believe, should deal primarily with the "automatic" processes and only very little with these other more indefinite ones. We would differentiate these three types of outcome by saying that the immediate specific and generalized outcomes are subject to measurement, and that for our purposes the intangible outcomes are not.

The foregoing statement, made in the preliminary report, was doubtless misunderstood by many teachers of algebra. The commonly accepted inference was that "standardization," according to our aims, methods, and criteria, meant complete "mechanization" of first-year algebra instruction. Nothing is farther from the truth concerning our real position than that. For that reason we wish to emphasize herewith: *Our position—the fundamental aim of first-year algebra: to develop ability to use algebraic methods in the solution of "original" or "novel" problems stated in verbal form.* Three years of research in this field has convinced us, however, that *efficiency in the solution of "original problems" is closely correlated with a thorough mastery of each of the tool operations.* This report will present detailed evidence on this point.

It has been shown that success in teaching algebra depends primarily on the teacher's knowledge of the typical difficulties which the pupils will meet in learning algebra. These difficulties may be found by testing the pupils' ability to manipulate the different formal operations and to use these in the solution of "translation" problems. To design tests for these operations necessitates a clear-cut statement and classification of the formal operations themselves. The committee now wishes to go on record as stating its belief that the *fundamental formal operations* of first-year algebra, ability in which should be tested, are those given in the following pages and that *ability in the manipulation of each of these should be made absolutely automatic.*

The necessity for automatic efficiency in the formal processes.—Teachers agree: that pupils must have *automatic* skill in manipulating the "tables" in arithmetic; that the spelling of "common words" shall be absolutely mastered (*automatized*) so that pupils will *never* make a mistake in spelling them; that a certain quality of handwriting shall be written by our pupils at a definite speed and that it shall be done *automatically;* that pupils shall perfect certain "*automatic*" *habits* of "reading" in the early years so that lessons in history and geography and literature shall not be turned into lessons in the development of the formal skill in getting meaning from the printed page.

In the same way, in order that the pupil may use successfully algebraic methods in the solution of verbally stated problems, he must have absolute mastery of the tool operations he is going to use in that solution. A pupil should remove parentheses, factor, solve simple equations, use special products, exponents, radicals, etc., just as he uses the multiplication table, writes, spells, or gets meaning from written language—*in a word, automatically.* It is not economic or expedient to force pupils to raise to "thinking" or "reasoning" levels, the formal manipulation of these purely tool operations. The Committee on Standards is, therefore, insisting on thoroughness in the formal operations, in the interest of "*economy of time*" *in first-year algebra; in other words, in order that a larger amount of time may be spent in the use of the formal operations in solving "original" problems.* This report will show methods by which, it is believed, this may be done.

The reader should be cautioned that "automatism" in the more complex processes (e.g., fractional equations) does not necessarily imply the instantaneous reaction of the pupil with the completely worked-out "answer" to the problem. In the case of problems containing but one or two steps the automatic response should be the "answer." But, in the cases of problems involving several steps, *automatism means the continuous unbroken reaction of the pupil with the proper steps in the solution.* The steps in the procedure of manipulation of the operation in question should be made completely a part of his automatic system of habits.

II. THE CONTENT OF FIRST-YEAR ALGEBRA

Our program next demands a complete classification of the content of first-year algebra. We have differentiated sharply the content of algebra between the formal processes of algebra and their application in verbal problems. Our preliminary investigation has shown that to test adequately the algebraic abilities of pupils in the use of algebraic symbolism in "original situations" *we must draw a sharp line between abilities in the formal processes and abilities in their application.*[1] The former may be measured by one type of test, *a formal time test,* the latter by another type, *the translation verbal test,* in which it is questionable whether there should be an "active" time limit. We give next a mere list of:

a) *The fundamental operations of first-year algebra ability in which should be tested.*—

1. Removal of parentheses
2. Combining terms
3. Subtraction
4. Evaluation
5. Special products
6. Factoring
7. Exponents
8. Clearing of fractions and fractional equations
9. Quadratic equations
10. Graphing of equations
11. Solution of "practical" formulae
12. Simultaneous equations

Mr. E. C. Denny, of the Illinois Committee on Standards, has made a very complete analysis of the content of first-year algebra as shown by currently used textbooks and the distribution of problems that indicate the types of opportunity for training which are now offered. It is expected that this report will be published separately; therefore the material will not be duplicated here. We may say in brief that the material collected in this phase of the work of the committee has been used to check up the organization of content and the interpretation of the results of testing pupils. His data lend added weight to our conclusions.

---

[1] See *School Review,* January, 1916, preliminary report of the Committee on Standards in First-Year Algebra.

b) *The types of verbal problems represented by our study.*—Verbal problems may be classified on two bases: (1) on the basis of degree of equations involved and the number of unknowns; (2) on the basis of the subject-matter composing the problems. The latter basis lends itself more easily to the design and classification of verbal problems for purposes of testing and was used in this investigation. The problems for which "scoring weights" have been worked out in this study fall under the following heads: (1) geometry problems (area, perimeter, angle, line, and volume problems); (2) physics problems (temperature, lever, falling bodies, formulae for sound, light, etc.); (3) motion problems; (4) problems involving ratio and proportion; (5) coin problems; (6) work problems; (7) mixture problems; (8) digit and number problems; (9) age problems; (10) clock problems; (11) percentage problems.

III. THE STANDARDIZED TESTS IN FIRST-YEAR ALGEBRA

The statement of the aims and outcomes of instruction and the content of the subject-matter paves the way for the design of tests for each of the fundamental operations. Our report[1] for 1915 treated fully the important question of *the determination of a valid method for measuring efficiency in the formal processes and in their application in verbal problems.* It was shown there (1) that the teacher-judgment method did not lead to an accurate determination of the difficulty of algebra problems; (2) that "mixed scales," containing both formal and translation problems, are not valid measures of the single types of mental processes involved in particular kinds of algebraic operations; (3) that the best hypothesis concerning difficulty of algebra problems is that based upon the proportion of a large group of pupils solving the problems.

The third revision of the Standardized Algebra Tests (Rugg and Clark) was submitted to teachers of algebra during the spring of 1916.[2] The tests have been constructed, criticized, and revised three times in accordance with the principles of design set forth last year. It should be pointed out that other algebra tests which have been made up, printed, and used by many school systems, do not make use of detailed principles of design. This investigation has convinced us that the following principles are essential to sound testing. We repeat the fundamental ones from last year's report:

---

[1] *School Review*, January, 1916.

[2] The eleven printed tests for the formal processes, together with complete lists of standardized verbal "translation" problems, may be secured in quantities from H. O. Rugg, School of Education, University of Chicago. The tests will be sold during this year at 4 cents per set (a set to a pupil). The fourth reprinting (within the next year) will enable us to sell these at a lower rate. The price will be set as nearly as possible on a "cost" basis, no profit being made on the tests.

a) PRINCIPLES GOVERNING THE SELECTION OF TEST PROBLEMS AND CONDUCT OF TESTS

In order that a scale for measuring efficiency in elementary algebra may successfully test automatic efficiency and independent solution it must be composed of two general types of test: (1) a specific test series (A) which will test the specific manipulative abilities of students in all the basic automatic operations involved in the solution of algebra problems; (2) a translation test (B) which will test the independent ability of the student in practical or applied problems.

1. *The formal tests*, as designed and presented herewith, conform to the following requirements:

a) They are made up of a series of problem-tests each of which is designed as a specific test for a definite automatic operation in algebra solution.

b) Each specific test is made up of a number of problems (10 to 28) each of an elemental nature and involving the operation in question, and for each of which the degree of difficulty has been determined carefully by experimentation.

c) *The cycle principle of design.*—Because it is impossible to arrange separate tests for all kinds of operations involved (owing to lack of time in classroom handling, etc.), those problems which involved *closely related operations were grouped in one test and arranged rigidly in rotation.* Thus the student solving 20 problems may be compared with the one solving 10 problems.

d) Each test was designed as a time test, the time being so arranged (by experimentation) that no student could quite finish the test in the time given, but so that all could do a considerable number—otherwise the measure of efficiency would have been too coarse. Care was taken to see that all pupils started and stopped the test at the same instant.

e) The directions were all given orally by the experimenters so that differences in rate of reading and comprehending directions might not complicate results.

f) Test problems were of the alternative sort; i.e., they were designed to give either right or wrong answers—otherwise careful evaluation and weighing of answers would have been necessary.

Time tests of formal processes in any subject of study must be designed in accordance with:

The cycle principle of rotation of problems.—The various ways in which the symbols, letters, etc., may be arranged for a given type of operation *should appear in exact rotation* in the test. For example, in Test I there are six principal ways in which parentheses problems may be "arranged," i.e., considering the use of the + and − signs, (), letters, etc. These appear in Test I in such order that in the first, seventh, thirteenth, nineteenth, and twenty-fifth problems the signs and symbols occupy the same relative positions, i.e., the problems involve the same algebraic and mental processes. Our research shows that *it is of the utmost importance that this "cycle principle" be followed in the most rigorous fashion.* The importance of this point is indicated by the differences in difficulty that are revealed by Table I, which shows how a few failures to follow the principle *exactly* caused quite different percentages of failure on the part of pupils in solving the problems.

Attention should be called to the fact that the validity of other algebra tests which have been drawn up without regard to this principle should be seriously called in question. For example, the Indiana Algebra Tests, based upon the "Standard Research Tests" devised by W. S. Monroe, sent out from the University of Indiana, have been made up in such a way that the efficiency of pupils solving a given number of problems on any test cannot be validly compared with the efficiency of pupils solving half as many, a third as many, twice as many, etc. To illustrate this point we reproduce Test II of this series of tests:

1. $4(3x-4)=$
2. $-5x^3(4x-1)=$
3. $-7(2+3x)=$
4. $-5(-4+6x)=$
5. $3(-1+6x)=$
6. $-4a^2(8x+4a^5)=$
7. $-7(-5x+8)=$
8. $-8x(-3x-5a^3)=$
9. $6(2-4x)=$
10. $-x(5-6x)=$
11. $-3(9+x)=$
12. $-5(-7x+3)=$
13. $-5(-4x-6y)=$
14. $-6y^2(-9-7x)=$
15. $-4(x-2)=$

Certainly no definite principle of design controls the placing of problems in this test. Our results show that *this is an essential* *step that must be followed if we are to have sound criticism of school practice in these matters.* It can be suggested from the results of our investigation that these problems, many of which are several times as difficult as others (problem 1 compared to problems 2, 6, 8, or 14, for example), are not put together in such a way as to lead to comparable results in testing pupils.

The construction of tests to measure the abilities of pupils should be based upon the most scientifically worked-out research principles. Recognizing the urgent need for care in such work, we have checked the "cycle principle" of rotating problems in two different years of experimentation. The results are given in Table I for readers interested in this phase of the work. This table gives the percentage of all of those pupils (over 2,500 took the various tests) who attempted each problem in each test, and who failed to work the problem in question correctly. In this investigation it has not been possible to inquire in detail into the effect of "practice" in working the recurring problems of the cycle. We believe that it will operate in the more difficult tests to give a gradually decreasing percentage of failures on successive problems of a particular type. Table I has been made up and examined carefully for the purpose of discovering which problems in each test are not roughly equal in difficulty to corresponding problems in other cycles. A problem in which a distinctly larger or smaller percentage of failures is found has been replaced by another. Careful study of such problems, in almost all cases, has revealed peculiarities in construction, *or in scoring*, that cause a problem to be thrown out. For example, in problem 3 in Test VI the percentage of failures is 3.5 per cent as compared to from 30 per cent to 36 per cent in all other corresponding problems. Study shows that the problem is so constructed that we cannot determine by an inspection of the pupil's answer whether the mental process is correct or not. The particular error that students make in this problem is that due to "adding exponents instead of multiplying them"; $(n^2)^2$ gives the same answer regardless of the process and thus the work of the pupil cannot be diagnosed to find out whether he is "right" or not. This method of analysis has been applied to each problem in each test and the accompanying list (Table I) of corrections is given.

The tests will be reprinted this year with these corrections made. We believe that they will then thoroughly justify the title "Standardized Tests in First-Year Algebra," and that they may be used by teachers to check up specifically the ability of their pupils in the formal operations.

TABLE I

PERCENTAGE OF PUPILS FAILING TO WORK CORRECTLY CORRESPONDING PROBLEMS ON SUCCESSIVE CYCLES OF EACH TEST

Problems are arranged in columns so that the percentage of pupils failing to work "*corresponding*" problems of successive cycles appear in any one column. To determine relative difficulty of *successive problems in one cycle read horizontally across the table.* Problems which call for study and redesign are indicated by inclosing the percentages thus ( ).

	TEST I (REMOVAL OF PARENTHESES)						TEST II (SPECIAL PRODUCTS)			
No. of Cycle	First Prob. in Each Cycle	Second Prob.	Third Prob.	Fourth Prob.	Fifth Prob.	Sixth Prob.	First Prob. in Each Cycle	Second Prob.	Third Prob.	Fourth Prob.
1	2.5	2.5	9.0	6.5	9.5	8.0	28.1	29.1	30.1	19.4
2	3.2	1.1	7.5	4.4	8.7	(12.2)	27.6	18.8	26.3	14.6
3	2.2	3.4	8.5	6.7	7.6	7.3	32.4	17.4	22.3	11.2
4	4.2	3.0	(12.7)	9.2	4.3	5.2	25.4	(5.84)	19.9	12.5
5	5.6	3.3	3.7	2.2	5.3	3.3	22.4	13.1	28.2	(7.82)
6	......	......	......	......	......	......	27.4	(8.0)	22.8	(21.3)

	TEST III (SUBSTITUTION)					TEST IV (FACTORING)				
No. of Cycle	First Prob. in Each Cycle	Second Prob.	Third Prob.	Fourth Prob.	Fifth Prob.	First Prob. in Each Cycle	Second Prob.	Third Prob.	Fourth Prob.	Fifth Prob.
1	32.5	34.8	33.2	27.0	33.0	41.5	10.1	7.65	8.6	33.6
2	25.8	29.3	31.1	22.0	27.3	33.7	9.8	12.9	8.4	23.5
3	20.0	19.1	22.5	12.6	16.1	32.2	9.5	8.6	5.3	27.3
4	14.5	23.4	18.5	12.9	11.8	29.1	11.4	(23.6)	5.7	(53.9)
5	......	......	......	......	......	29.5	5.6	13.5	(11.1)	10.9

	TEST V (FRACTIONAL EQUATIONS)					TEST VI (EXPONENTS)					
No. of Cycle	First Prob. in Each Cycle	Second Prob.	Third Prob.	Fourth Prob.	Fifth Prob.	First Prob. inEach Cycle	Second Prob	Third Prob.	Fourth Prob.	Fifth Prob.	Sixth Prob.
1	63.0	44.0	53.1	52.6	(41.7)	8.5	(37.7)	(3.5)	11.6	(16.1)	3.5
2	61.7	36.9	58.8	(35.5)	(66.9)	8.5	8.5	32.7	14.1	(36.2)	2.5
3	55.4	33.1	61.8	48.5	55.0	4.0	7.1	33.9	11.7	45.9	6.2
4	65.2	33.3	(36.6)	57.2	51.6	3.7	10.7	29.2	10.9	38.9	(1.7)
5	59.8	31.3	(40.0)	52.6	31.6	7.9	8.1	33.5	9.2	43.1	2.1
6	......	......	......	......	......	3.4	9.2	36.0	14.8	46.2	6.4

TABLE I—*Continued*

TEST VII (QUADRATIC EQUATIONS)				TEST VIII (GRAPHING EQUATIONS)					
No. of Cycle	First Prob. in Each Cycle	Second Prob.	Third Prob.	First Prob.	Second Prob.	Third Prob.	Fourth Prob.	Fifth Prob.	Sixth Prob.
1......	19.7	34.3	23.7	55.9	53.5	43.4	51.8	50.5	46.4
2......	22.0	29.0	24.8	(Only one cycle)					
3......	17.6	28.5	40.4						
4......	20.6	29.1	43.8						
5......	15.6	29.9	35.7						
6......	13.9	(39.1)	33.0						
7......	10.7	30.7	25.8						

TEST IX (RADICALS)				TEST X (SOLUTION OF FORMULAE)						TEST XI (SIMULTANEOUS EQUATIONS)		
No. of Cycle	First Prob. in Each Cycle	Second Prob.	Third Prob.	First Prob. in Each Cycle	Second Prob.	Third Prob.	Fourth Prob.	Fifth Prob.	Sixth Prob.	First Prob. in Each Cycle	Second Prob.	Third Prob.
1.......	23.1	41.1	50.8	(34.2)	42.0	64.3	95.1	64.6	60.1	(32.5)	23.1	37.6
2.......	16.0	44.2	51.7	46.0	43.0	60.1	92.7	62.0	62.1	(17.9)	27.7	23.1
3.......	(35.5)	42.5	44.3	56.2	39.0	57.8	94.6	57.1	59.3	23.9	17.1	20.9
4.......	18.6	28.4	50.0	53.8	41.0	55.8	93.8	52.3	56.9	23.0	(10.0)	32.5
5.......	(33.1)	33.6	45.7							25.8	19.7	7.4
6.......	14.7	30.0	(59.8)									
7.......	8.2	(19.8)	48.7									

B. THE FIRST CYCLE OF EACH OF THE ELEVEN FORMAL TESTS TOGETHER WITH THE SPECIFIC CHANGES THAT WILL BE MADE IN REPRINTING THE TESTS IN 1917

We give next the exact problems used in the first cycle of each test. Teachers of algebra can thus see exactly how many different arrangements of material are offered in each specific operation. We also list the exact changes that are to be made in each problem. To understand this discussion thoroughly the reader should secure copies of the tests and compare each change with the data of Table I.

1. *Removal of parentheses.*—Tested by Test I. Time, 2 minutes.

First Cycle

$$6(3x+8)$$
$$5(4x-2)$$
$$-3(8x+3)$$
$$-4(3x-4)$$
$$9(-7x-1)$$
$$-8(-4x-7)$$

No changes to be made in particular problems in redesigning tests this year.

2. *Special products.*—Tested by Test II. Time, 3 minutes. Each cycle now contains 4 problems, arranged as follows in first cycle:

1. $(2x-3)^2$
2. $(3m+n^2)(3m-n^2)$
3. $(a-4)(a+5)$
4. $(2b+3a)(2b-3a)$

Because of the relative importance of the following type, No. 5, i.e., $(5x+1)(x+3)$, Test II will be redesigned to give a cycle in which the order of problems is as follows: problem 4 being replaced by No. 5. First cycle to read:

$(2x-3)^2$
$(3m+n^2)(3m-n^2)$
$(a-x)(a+5)$
$(5x+1)(x+3)$

Thus the fourth problem in each cycle will be the corresponding problem from Test IV (factoring), that is, Nos. 5, 10, 15, 20, 25.

3. *Substitution or evaluation.*—Tested by Test III. Time, 5 minutes.

First Cycle
If $x=4$ and $y=2$ what does $2x^2-3xy=$?
If $a=3$ and $b=2$ “ “ $3ab+ab^2=$?
If $c=2$ and $d=5$ “ “ $cd^2-2cd=$?
If $p=4$ and $q=3$ “ “ $p^2+4pq=$?
If $x=3$ and $y=5$ “ “ $x^2+2x^2y=$?

No changes to be made in this test.

4. *Factoring.*—Tested by Test IV. Time, 5 minutes.

First Cycle
$5x^2+15x^3$
a^2-64
y^2-6y+9
$b^2+11b+28$
$5x^2+16x+3$

Changes to be made in particular problems (18, 20, 24):

Prob. No.	Now reads:	It will read:
18	$p^2-12p+36$*	$p^2-18p+81$*
20	$9x^2+33x+18$*	$3x^2+22x+35$*
24	$a^2+10a+24$*	$a^2+10a+21$*

---

*$p^2-12p+36$ and $9x^2+33x+18$ offer many possibilities in factoring, e.g., 6, 6; 9, 4; 18, 2; 12, 3; 36, 1; more than one of which gives a middle term approximating 12. Corresponding problems in other cycles do not give this. Hence such problems will be changed as indicated above.

5. *Solution of fractional equations.*—Tested by Test V. Time, 12 minutes.

First Cycle

$$\frac{4x-2}{3}-\frac{x-3}{4}=0$$

$$\frac{x+1}{x-1}=\frac{5}{3}$$

$$\frac{4}{3+x}-\frac{2}{1+x}=0$$

$$\frac{x+8}{x+9}=\frac{x-5}{x-7}$$

$$\frac{3x}{4}+\frac{3x}{2}=5$$

Changes to be made in particular problems in the test (9, 10):

Prob. No.	Now reads:	It will read:
9	$\frac{x-6}{x+2}=\frac{x-4}{x+5}$*	$\frac{x-6}{x-2}=\frac{x+4}{x+5}$*
10	$\frac{x}{24}+\frac{8x}{4}=17$†	$\frac{x}{12}+\frac{8x}{4}=2$

---

* In dividing by the coefficient of x this problem did not involve the use of signs, all other corresponding problems did.

† Change made because of unequal degree of arithmetic difficulty.

6. *Exponents.*—Tested by Test VI. Time, 3 minutes.

First Cycle

$$a^3\cdot a^5=$$

$$5x^7\cdot 6a^5=$$

$$(n^2)^2=$$

$$\frac{c^3}{c^2}=$$

$$(ba^2)^2=$$

$$x^7\cdot x=$$

Changes (2, 3, 5):

Prob. No.	Now reads:	It will read:
2	$5x^7\cdot 6a^5$*	$5x^7\cdot 6x^5$*
3	$(n^2)^2$†	$(n^2)^3$†
5	$(ba^2)^2$†	$(ba^3)^4$†

---

* Mistake in printing.

† Answer does not reveal specific operation.

7. *Solution of quadratic equations.*—Tested by Test VII. Time, 7 minutes.

First Cycle

$$x^2-81=0$$

$$y^2+y=6$$

$$n^2-7n=-12$$

Changes (9, 12, 17):

Prob. No.	Now reads:	Will read:
9	$m^2+5m=6$	$m^2-5m=14$
12	$q^2+13q=30$	$q^2-11q=26$
17	$a^2-10a=24$*	$a^2-6a=27$*

*Too many possibilities of factoring the 24.

Additions.—No recurring problems of the type $m^2+5m=6$ were used in the 1916 printing of the tests. They will be reprinted to include this type as an additional problem in each cycle.

8. *Finding roots by graphs.*—Tested by Test VIII. Time, 12 minutes.

Six graphing problems on the sheet of this type:

Find the roots of $x-y=4$

$2x+y=5$ by drawing the graphs.

No changes.

9. *Expressing radicals in simplest radical form.*—Tested by Test IX. Time, 3 minutes.

First Cycle

$\sqrt{8}$

$\sqrt{a^3b^4}$

$\sqrt{2/3}$

Changes (6, 7, 13, 18, 19, 20):

Prob. No.	Now reads:	Will read:
6	$\sqrt{3/5}$*	$\sqrt{3/10}$*
7	$\sqrt{32}$	$\sqrt{12}$
13	$\sqrt{48}$	$\sqrt{45}$
18	$\sqrt{3/8}$†	$\sqrt{3/11}$
19	$\sqrt{20}$	$\sqrt{24}$
20	$\sqrt{b^3c^2}$	$\sqrt{b^7c^{10}}$

* To avoid duplication.

† Multiplying by denominator introduces in numerator perfect square factor.

10. *Solution of simple equations (practical formulae).*—Tested by Test X. Time, 7 minutes.

First Cycle

$$V=LWh$$

$$c=\frac{E}{R}$$

$$E=\frac{PL}{K}$$

$$L=\frac{Mt-g}{t}$$

$$I=\frac{bd^3}{3}$$

$$E^2=\frac{JWhr}{t}$$

Changes (1, 19):

Prob. No.	Now reads:	Will read:
1	$V=LWh$	Interchange No. 1 and No. 19 because of common use of No. 1.
19	$P=ahw$	

11. *Simultaneous equations.*—Tested by Test XI. Time, 12 minutes.

$$\begin{aligned} 2x+y &= 10 \\ 3x-2y &= 1 \end{aligned}$$

$$\begin{aligned} 2x &= 3y+3 \\ 5x+3y &= 39 \end{aligned}$$

$$\begin{aligned} 4n-2r &= 0 \\ 3n+5r &= 13 \end{aligned}$$

Problems in which changes will be made in redesigning Test XI (4):

Prob. No.	Now reads:	Will read:
4	$5b+c=21$ $6b-2c=22$	$5b+c=21$ $3b-2c=10$

ADDITIONAL "FUNDAMENTAL OPERATIONS" FOR WHICH TESTS WILL BE DESIGNED THIS YEAR

1. *Subtraction.*

   Sample cycle:

   1st problem: From $2a-3b+5c$ take $5a+4b-c$
   2d problem: Take $4x+2y-6z$ from $2x-y+z$
   3d problem: Subtract $3r-5x+10t$ from $r+s+4t$

2. *Clearing of fractions.*
3. *Combining terms.*

C. METHODS OF MEASURING ABILITY TO USE ALGEBRAIC METHODS IN THE SOLUTION OF "ORIGINAL" VERBAL PROBLEMS

Algebraic abilities are of two distinct kinds: (1) that involved in translating "word problems" into algebraic symbolism; (2) that involved in manipulating the formal operations. In the testing done in 1915 by this committee the former type (involving "reasoning ability") was measured by tests which required the pupil to do two distinctly different things: (1) state the verbal *problem* in algebraic form (i.e., give the equational relation); (2) work through the manipulation of the formal operations to the statement of the final answer. The result was that the incorrectness of the answer of the pupils did not indicate which type of error had been made. The revised method of testing "reasoning" abilities is to require the pupil to do but one thing: *translate the problem into algebraic language.* The ability to manipulate the formal operation is measured by the appropriate formal test. In this way our scoring is *not* complicated by two or more different types of mental process.

THE DESIGN OF "REASONING TESTS"

a) Shall they be time tests?—In our 1915 principles of design, we established the principle that efficiency of manipulation of formal operations should be measured by the number of problems of a given type that could be done in a unit of time—that is, *formal tests must be time tests.* On the other hand, since we have as yet in educational research no definite means of equating speed and accuracy in working problems, we have decided to determine the difficulty of "verbal problems" (and thus the "weights" to be used in scoring pupils) by having pupils solve them without the complicating pressure of time.

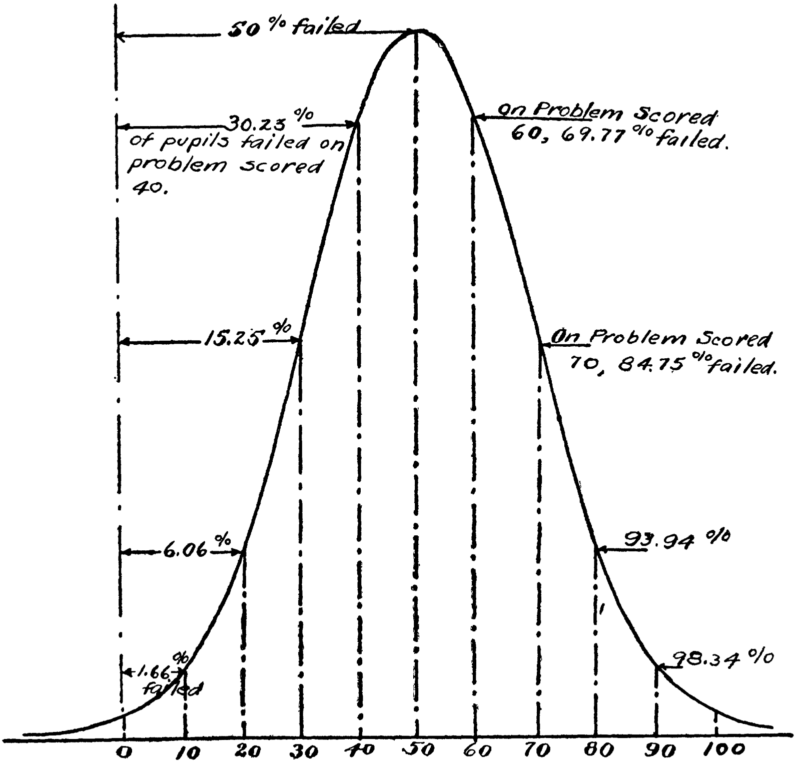
b) Principles of design of verbal tests.—Thus the principles of design underlying the two methods of measuring ability are: (1) for the formal operations, group together in one test all necessary arrangements of letters, symbols, etc. (observing rigidly the "cycle" principle), and measure the efficiency by the number of problems worked in a unit of time (say, a minute); (2) for the verbal problems: (*a*) design tests of verbal problems (a score or more of such lists ought to be available eventually) ranging in degree of difficulty from very easy problems (which nearly all pupils will solve correctly) to very difficult problems (which but few pupils will solve correctly); (*b*) weight each problem in scoring the ability of pupils by determining its relative degree of difficulty; (3) this can be done by (*a*) finding the percentage of a large and representative group of pupils which solve each problem correctly; (*b*) assuming that algebraic abilities are distributed in the general first-year high-school population in accordance with some known distribution curve.

Working on these hypotheses and principles, lists of verbal problems (totaling 51 in all) were drawn up covering the principal types of subject-matter named above. As a result of giving the 1915 tests, problems of widely varying degrees of difficulty were included. These problems were then worked by 1,295 pupils, distributed throughout 26 school systems, 17 of which also worked the 11 formal tests. As a result of this testing there was determined the percentage of the group that worked each problem correctly. In order then to determine the relative difficulty of each problem, the assumption was made that algebraic ability is distributed fairly closely in accordance with the "normal" probability curve. (Intellectual abilities in the elementary school have been shown to follow this distribution rather closely. We recognize the possible existence of many factors which tend to make the secondary-school curve skewed to the high end of the scale. Almost nothing is actually known of the amount and direction of their influences, however. The best "practical guess" that can be made at the present time as to the distribution of scholastic abilities is that it corresponds closely enough to the curve of error to warrant using the well-worked-out properties of that curve in our design.)

Let Diagram I represent the distribution of algebraic abilities in the pupils represented by our 27 school systems. The base line then represents a "scale of algebraic difficulty" ranging, let us say, from nearly 0 ability to nearly perfect or 100 per cent ability. The area between the curve and the base line represents the number of pupils in our entire group. If we divide the base line into any number of parts and erect upright lines at the points representing these parts we could determine, from the properties of the normal curve, the number of pupils that ought to be found between these distances in the base line.

In the same way we could determine what percentage of our group of pupils should be found distributed between the zero point on the base line and any other point. Since the normal curve has the property that it actually meets the base line only at infinity we are forced to set our 0 and 100 points arbitrarily by deciding how large a percentage of the entire group we may drop off at both ends of the base line.

Taking as our unit of measurement on the base line, sigma, the "standard deviation" of the distribution (indicated graphically on Diagram I), and laying it off 2.5 times each way from the midpoint of the curve, gives us 5 divisions (which may be conveniently divided into 10 divisions corresponding "practically" to our public-school marking system). In doing this we are arbitrary to the extent of neglecting only 0.62 of 1 per cent of our pupils at each end of the base line. If this 0.62 of 1 per cent is thrown into the middle of the curve where the individuals are more closely grouped, it is a negligible factor. Calling the point 2.5×sigma from the mid-point 0, and setting the successive points 10, 20, 30, etc., to 100, we now have a practical working "scale of algebraic difficulty"



Scale of Algebraic Difficulty

Diagram 1.—Distances on the base line represent, to scale, relative difficulty of problems. Area under the curve represents total number of pupils that were tested for ability to translate verbal problems. 0 and 100 points set arbitrarily at $2.5\times\sigma$ from the mean. Mean is set arbitrarily at 50. Area of the curve between 0 and any point on base line represents percentage of pupils who failed the problem placed at that point.

over the successive points of which the corresponding percentages of our pupils may be indicated. Doing this, we see in Diagram I the proportions of our group of pupils that correspond to various degrees of difficulty on the base line. Thus a problem which is failed by 96.6 per cent of the group falls at the point marked 85; that failed by 84.8 per cent is scored 70, etc., throughout the list. To enable us to mark in an accurate way, a table has been computed in which the base line has been divided into 500 parts.

Verbal list A is printed herewith as representative of the scoring of the 51 problems that have been worked by our entire group of pupils. Each of the 51 problems has been scored for difficulty.[1]

The problems in each list range from very easy to very difficult by approximately equal steps. They are scored on an absolute percentile scale from 25 to 85. The outstanding deficiency is found to be: no problems were included in our list of 51 harder than 85 on this scale (assuming a normal distribution of ability) or easier than 25. The problems at the low end of the scale were estimated incorrectly to be so easy to solve that all pupils would be able to work them. It was found that no problem showed less than 10 per cent of failures and no problem showed less than 3 per cent of successes. The committee is now initiating a thoroughgoing attempt to have large numbers of pupils solve several hundred problems. From these solutions we shall be able to extend the present list and add many more.

When the standardizing of verbal problems is complete, teachers may use such material very easily in a thoroughly practical way. For example, a teacher may, at any time make up an "examination" of translation problems by selecting one or more problems from each "weight," 25, 30, 35, etc., or from "weights" particular degrees of difficulty. She can facilitate her own testing by selecting problems in such a way that the total maximum score is, say, 500 or 1,000—easily reduced to a percentage basis. It must be remembered in all this work that we are testing here *the results of one year's training in algebra*. There must be further detailed experimentation before the comparability of scores obtained *during* the year with those obtained at the end of the year will be established.

---

[1] Complete lists of verbal problems may be secured from H. O. Rugg, School of Education, the University of Chicago.

STANDARDIZED TESTS IN FIRST-YEAR ALGEBRA (VERBAL PROBLEMS)

Rugg and Clark

(Graded on an absolute scale of difficulty from 0 to 100, in accordance with two assumptions: (1) among first-year high-school students the distribution of algebraic ability approximates the "normal" probability curve; (2) the difficulty of problems varies as the percentage of pupils able to solve the problems correctly, superimposed on the base line of this curve.)

Score	VERBAL LIST A
85	A train running from Chicago to Denver at average speed of 40 miles an hour, takes 3 hours longer to make the run than one running at 45 miles an hour. What is the distance from Chicago to Denver?
80	If a cistern can be filled by one pipe in x minutes and emptied by another in $x+5$ minutes, what part of the cisternful runs in one minute if both pipes are open?
75	Find two numbers whose sum is 51, such that if the greater is divided by their difference, the quotient is $3\frac{1}{2}$.
70	Twice the width of the Pennsylvania station in New York exceeds its length by 80 feet. 4 times the length exceeds the perimeter by 700 feet. Find the dimensions.
65	If a boy $4\frac{3}{4}$ feet tall casts a shadow $4\frac{1}{2}$ feet long at the same time that a school building casts a shadow $57\frac{1}{2}$ feet long, how high is the school building?
60	A father 54 years old has a son aged 9 years. In how many years will the age of the father be just 4 times that of the son?
55	Two boys play at teeter. One weighs 100 lbs. and sits 6 feet from the point of support. The other weighs 120 lbs. How far from the point of support must he sit in order to make the board balance?
50	What number has the property that when multiplied by $\frac{5}{6}$, the result is greater by 1 than when multiplied by $\frac{4}{5}$?
45	If the width of a rectangle is W increased by 10 and its length L increased by 20, write the equation for its perimeter.
40	8 times a certain number equals 45 diminished by the number. State the equation by which you would find the number.
35	If W and L are the width and length of a rectangle, write the equation for its area in terms of W and L.
30	If you represent a number by x, how will you represent 5 more than 5 times the number?
25	Express the following verbal statement in algebraic form: the square of a side plus five.

[*To be continued in March*]